QUALITY OF DATA DOWNLOADS

BIBLIOGRAPHIC DATABASES FREQUENTLY USED FOR SYSTEMATIC REVIEWS

RACHAEL POSEY
Systematic Review Coordinator & Librarian to the Eshelman School of Pharmacy
Health Sciences Library, UNC – Chapel Hill

JENNIFER WALKER
Cancer Information Librarian
Health Sciences Library, UNC – Chapel Hill

Examine quality of data exported to…

Prioritize order records are uploaded into EndNote to retain the primary record with best quality of data

---

Other implications & applications…

➢ Identify problem areas for duplicate removal from other citation managers

➢ Decide between databases with similar coverage which one has better quality data
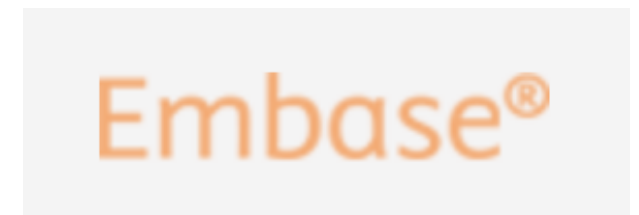
METHODS

random sample
277
journal articles

→

Records
2770
scored

SAMPLE SIZE

Downloaded records from major bibliographic databases

4 databases

2 download options

Embase	.RIS / Plain Text
Ovid Medline	.RIS / CGI
PubMed Medline	.NBIB / Plain Text
Google Scholar	Mendeley / 1 each

PRESENT

COMPLETE

ACCURATE

PRESENT

📖

Information in the field was present and in the field it was supposed to be in

COMPLETE

All information was complete

Examples of incompleteness:
- Titles cut off half way through
- Only first page was given rather than page range
- Only first few authors were listed instead of all

ACCURATE

Information brought in was correct (had no mistakes) & no additional information

Examples of inaccuracies:
Author fields sometimes included affiliations or degree abbreviations

Fields scored for presence, completeness & accuracy

Journal name

Article title

Author

Abstract

Volume

Issue

Year

URL

Accession

DOI

SPECIAL CASES

Abbreviations

Journal Titles

Health Services Research

⬇

Health Serv Res

Page Numbers

446 – 452

⬇

446 – 52

SPECIAL CASES
URLs

Multiple URLs

Database record (with proxy)

Database record (no proxy)
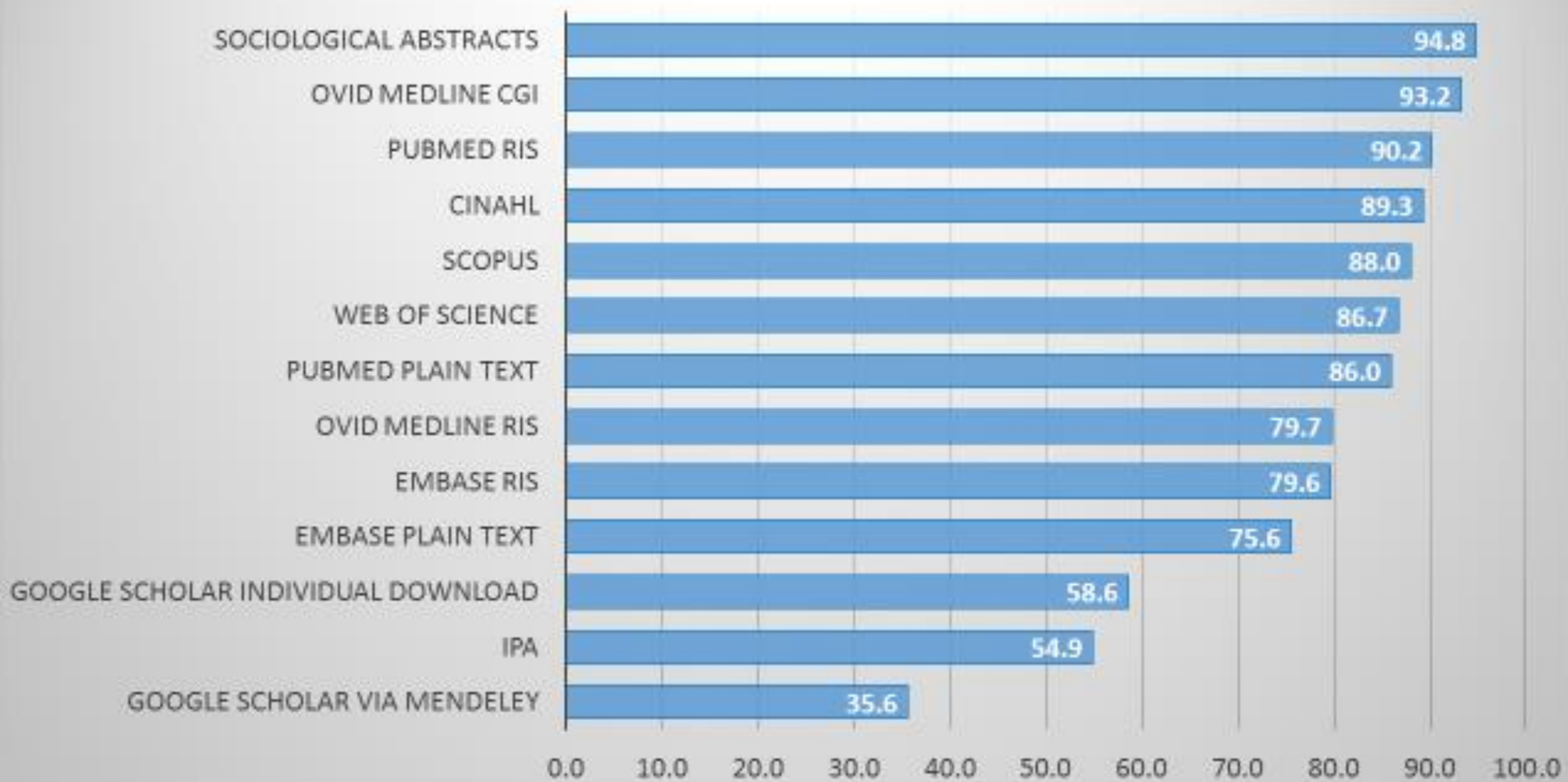
Accession Number

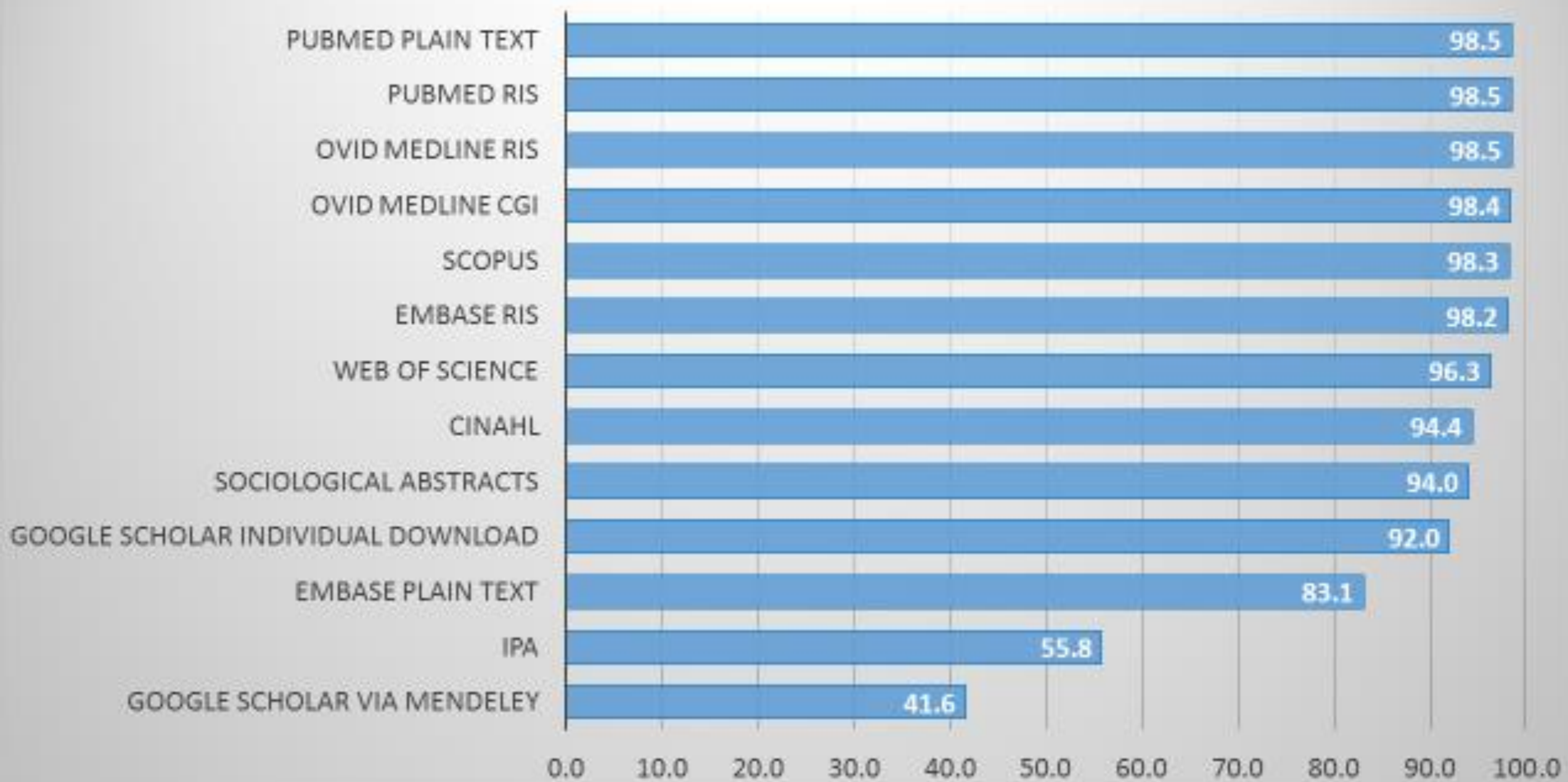Journal record (no proxy)

ANALYSES

- Overall average score by database for **all fields**

- Overall average score by database for **citation information only**

- Overall scores for each database for **other fields (URL, DOI, Abstract)**

- **Head-to-head scores** for each database

RESULTS
&
DISCUSSION

Overall Mean Score for All Fields

	Score
SOCIOLOGICAL ABSTRACTS	94.8
OVID MEDLINE CGI	93.2
PUBMED RIS	90.2
CINAHL	89.3
SCOPUS	88.0
WEB OF SCIENCE	86.7
PUBMED PLAIN TEXT	86.0
OVID MEDLINE RIS	79.7
EMBASE RIS	79.6
EMBASE PLAIN TEXT	75.6
GOOGLE SCHOLAR INDIVIDUAL DOWNLOAD	58.6
IPA	54.9
GOOGLE SCHOLAR VIA MENDELEY	35.6

Overall Score, All Citation Information

Source	Score
PUBMED PLAIN TEXT	98.5
PUBMED RIS	98.5
OVID MEDLINE RIS	98.5
OVID MEDLINE CGI	98.4
SCOPUS	98.3
EMBASE RIS	98.2
WEB OF SCIENCE	96.3
CINAHL	94.4
SOCIOLOGICAL ABSTRACTS	94.0
GOOGLE SCHOLAR INDIVIDUAL DOWNLOAD	92.0
EMBASE PLAIN TEXT	83.1
IPA	55.8
GOOGLE SCHOLAR VIA MENDELEY	41.6

Head-to-Head Comparisons, Overall Score by Db

Database	Score
OVID MEDLINE CGI	17.1
PUBMED RIS	14.0
CINAHL	12.0
SCOPUS	11.3
SOCIOLOGICAL ABS	11.1
PUBMED PLAIN TEXT	9.2
WEB OF SCIENCE	8.9
OVID MEDLINE RIS	2.6
EMBASE RIS	0.6
EMBASE PLAIN TEXT	-1.6
IPA	-19.6
GS INDIVID DL	-22.5
GS VIA MENDELEY	-44.0
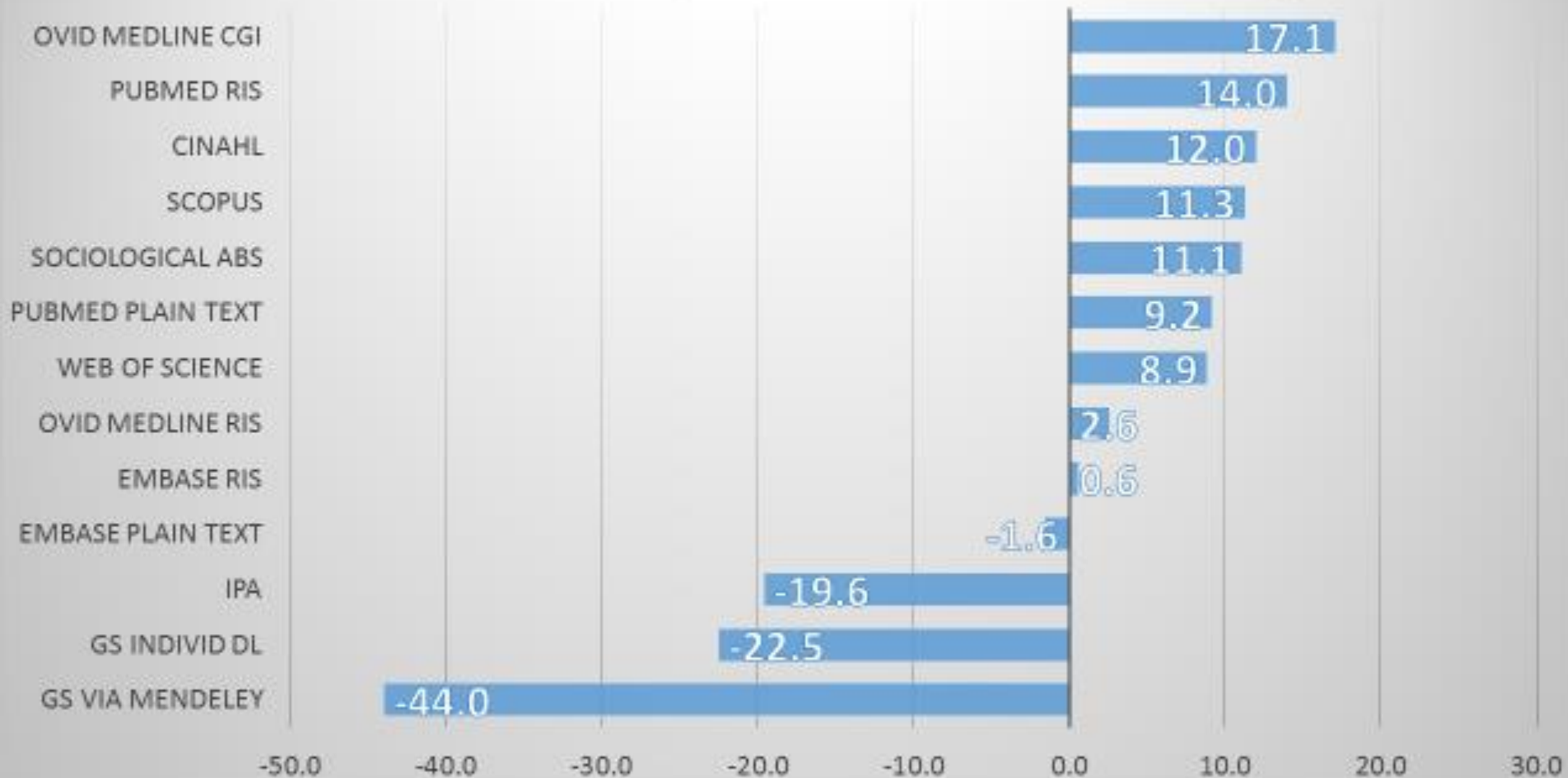
Head-to-Head Comparisons, Citation Score by Db

Database	Score
OVID MEDLINE CGI	11.7
PUBMED RIS	11.1
PUBMED PLAIN TEXT	11.0
SCOPUS	10.4
OVID MEDLINE RIS	10.3
EMBASE RIS	10.2
CINAHL	8.4
WEB OF SCIENCE	8.3
GS INDIVID DL	4.2
SOCIOLOGICAL ABS	2.7
EMBASE PLAIN TEXT	-6.1
IPA	-33.8
GS VIA MENDELEY	-51.0
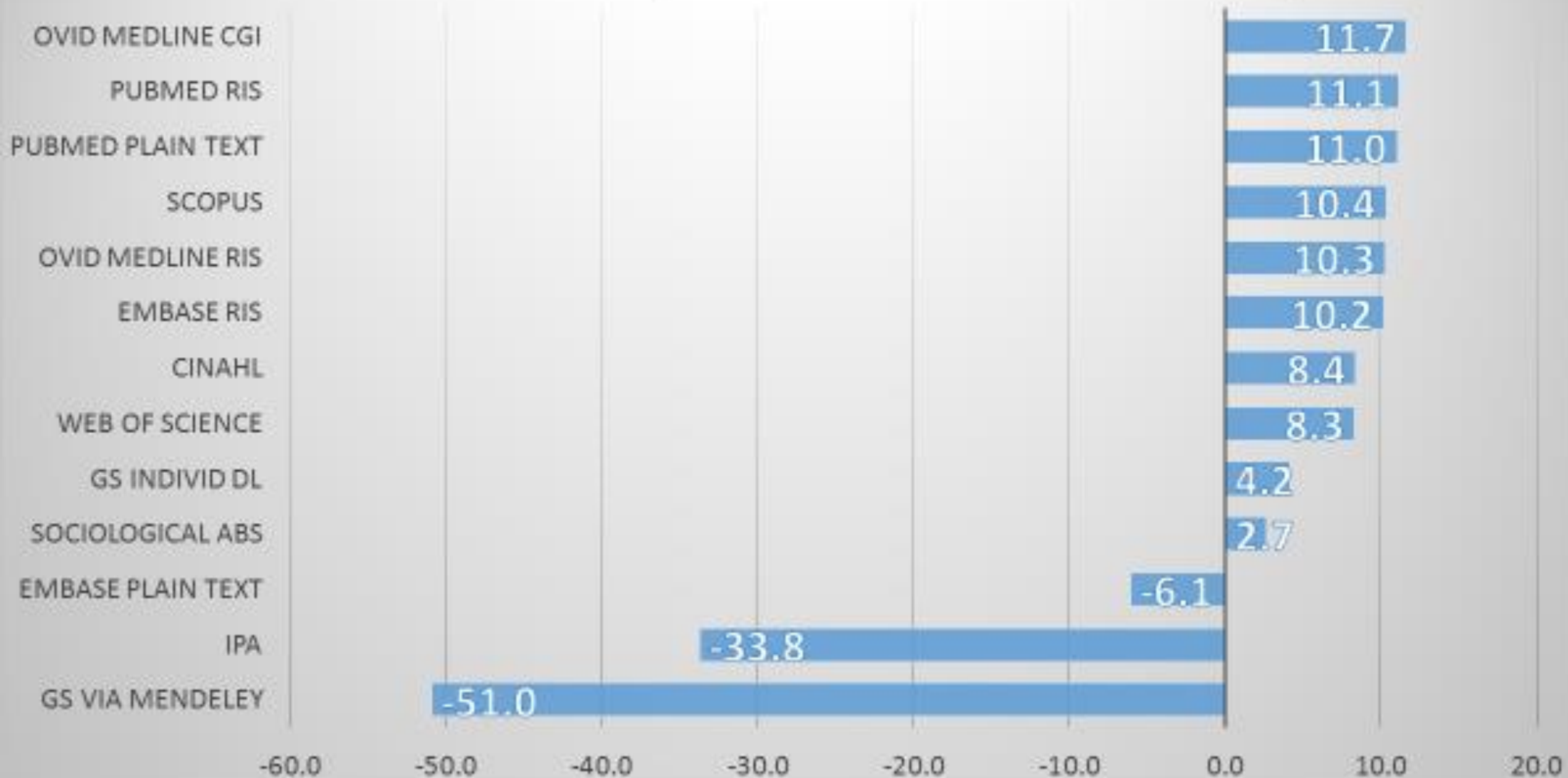
CONCLUSIONS

CONCLUSIONS
Database Order for EndNote

- Ovid MEDLINE CGI or PubMed NBIB **1st**

- CINAHL, Scopus, Sociological Abstracts, WoS **2nd**

- Embase RIS, IPA, and Google Scholar **3rd**

(Title or Title/Year matches best option)

CONCLUSIONS
Other Citation Managers

- DOIs inconsistent

- Double-check smaller databases

- Double-check Google Scholar

CONCLUSIONS
Databases with Similar Coverage

PubMed and Ovid MEDLINE scored roughly the same, and better than Embase

Scopus and WoS scored roughly the same and better than Google Scholar

Smaller databases have wide variances in data quality and the fields they bring in

CONCLUSIONS
Format Selections

Ovid CGI over RIS

PubMed NBIB over plain text

Embase RIS over plain text

Google Scholar individual download over Mendeley/Zotero/F1000 browser plugins

ACKNOWLEDGEMENTS

Richard Faldowski, PhD
Research Methodologist and Associate Professor
Department of Allied Health Sciences, School of Medicine
University of North Carolina at Chapel Hill

QUESTIONS

Rachael Posey
rposey@email.unc.edu

Jennifer Walker
jennifer_walker@unc.edu